DOCUMENT RESUME

ED 358 156 TM 019 955

AUTHOR O'Sullivan, Rita G.; Johnson, Robert L.
TITLE Using Performance Assessments To Measure Teachers' Competence in Classroom Assessment.
PUB DATE Apr 93
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary School Teachers; Elementary Secondary Education; *Graduate Students; Higher Education; Measurement Techniques; *Methods Courses; Pilot Projects; *Questionnaires; Secondary School Teachers; Student Evaluation; *Teacher Competencies; Teacher Education; Teacher Evaluation; Teaching Methods; Thinking Skills; Transfer of Training
IDENTIFIERS *Assessment Literacy; *Performance Based Evaluation; Standards for Teacher Competence (AFT NCME NEA); Teacher Competencies Assessment Questionnaire

ABSTRACT

The development and pilot testing of a set of performance assessments to determine classroom teachers' measurement competencies in areas covered by "Standards for Teacher Competence in Educational Assessment of Students" (1990) are described. How the use of performance assessments in a graduate-level classroom-assessment course can enhance student learning and increase the transfer of the measurement concepts to the classroom is shown. Twenty-nine graduate students enrolled in a course in educational measurement were pretested and posttested with the Teacher Competencies Assessment Questionnaire (TCAQ) during a classroom assessment course. Eight performance-based tasks, matched with the "Standards," were developed for the class. Results indicate that students who complete a measurement course that is performance-based increase their assessment competence as measured by the TCAQ. Performance-based educational measurement courses, in emphasizing higher order thinking skills, actively involving students in the learning process, and personalizing instruction, may increase the relevance of assessment to practitioners, and may increase the likelihood that measurement concepts will carry through to the classroom. Three tables present study data. Attachments include a list of assessment tasks for individuals and partners, a course outline, and a list of classroom assessment tasks. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED358156

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

RITA O'SULLIVAN
ROBERT JOHNSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Using Performance Assessments to Measure Teachers' Competence in Classroom Assessment

Rita G. O'Sullivan & Robert L. Johnson

School of Education
University of North Carolina at Greensboro
Greensboro, NC 27412
919/334-4095

April 1993

Paper presented at the annual meeting of the American Educational Research Association, Atlanta, Georgia.

The authors would like to acknowledge the assistance of the School of Education and the Collegium for Schools, Schooling, and Education at the University of North Carolina at Greensboro in presenting this paper.

TM019955

Using Performance Assessments to Measure Teachers' Competence in Classroom Assessment

Rita G. O'Sullivan & Robert L. Johnson
School of Education
University of North Carolina at Greensboro

Educational Importance

Teachers spend much of their time assessing students' progress, and yet most teacher education programs and state teachers certification agencies do not require pre-service teachers to complete any special training in educational measurement (O'Sullivan & Chalnick, 1991; Schafer & Lissitz, 1987) Recently, Marso and Pigge (1992) reviewed 225 studies that focused on teachers' competence in test development to find that inadequacies identified 25 years ago persist. Compounding this problem is the persuasive argument that teachers are becoming more, not less, important to the assessment process in the context of current educational reform movement (Jett & Schafer, 1992; Wolf, LeMahieu, Eresh, 1992) . Not surprisingly, issues concerning classroom assessment are of growing concern among educators.

Teachers' ability to effectively measure students' progress is central to this growing concern. In response, the National Council of Measurement in

Education (NCME), the American Federation of Teachers (AFT), and the National Education Association (NEA) jointly produced the Standards for Teacher Competence in Educational Assessment of Students (AFT/NCME/NEA, 1990). This document contains seven standards of assessment knowledge or skills that teachers are expected to possess. The joint concern of these organizations and others continues through the development of standards for public school administrators' assessment competence (Plake, Impara, & Sanders, 1992) and through a national survey of educators' current level of assessment competence (Plake, Impara, & Fager, 1992).

While one focus has been on setting standards of what measurement competencies teachers should possess, another focus has been on changing the nature of classroom assessment. Curriculum reformers have joined measurement specialists in a nationwide discussion of alternative classroom assessment procedures consistent with new curriculum approaches. Performance assessments, alternative assessments, and authentic assessments have been frequent subjects of recent discussion in the measurement, curriculum, and research literature (e.g., Frechtling, 1991; Linn, Baker, Dunbar, 1991; Mehrens, 1992; Moss, P. A., Beck, J. S., Ebbs, C., Matson, B., Muchmore, J., Steele, D., Taylor, C., & Herter, R., 1992; O'Neil, 1992; Shavelson, Baxter, & Pine, 1992; Wiggins, 1989). In general there is much support for broadening the type of assessments used in classrooms but much concern that these assessments be carefully constructed and used appropriately.

Gong, Venezky, and Mioduser (1992) reported how the use of instructional assessment, integrating performance assessment with instructional goals, promoted positive curriculum reform in science education. They clearly emphasized the importance of training teachers to use instructional assessment

so that fundamental, systemic, positive change in education can occur. Further they argued that teachers need to model effective learning processes so that students see how they may become active, productive learners.

Although there is appreciable support for training teachers in measurement (including alternative assessment procedures) and some discussion about how such training should occur (Airasian, 1991; Stiggins, 1991a; Wiggins, 1992; Wise, Lukin, & Roos, 1991) there are few empirical studies reporting the results of such training. O'Sullivan, McColskey, and Harman (1992) reported the results of a program to train middle-grades science teachers in the use of performance assessment in connection with a National Science Foundation-funded curriculum that was being developed. They conducted the training by having teachers develop performance assessment for the new sixth-grade science curriculum. They found that most of the 15 teachers they worked with reacted positively to using performance assessment approaches but that the process was gradual and that teachers experienced difficulty breaking away from traditional content knowledge assessment practices.

These events throughout the country are important, but even when teachers receive formal assessment training, there is no guarantee that they will find it useful (Stiggins, 1991b). Often educational measurement courses lack relevance to practitioners. Possibly the same criticism leveled at classroom teachers for not stressing higher order thinking skills, actively involving students in the learning process, or personalizing instruction could be aimed at college instructors of educational measurement.

Plake and Impara (1992) encourage researchers to use the national survey Teacher Competencies Assessment Questionnaire for designing classroom assessment training for teachers. They caution, however, that the

reliability estimate from the national sample (r equal to approximately .60) is not sufficient for decisions to be made about the competencies of individuals. Performance assessments are more appropriate to determining the classroom assessment competencies of individual teachers. Performance assessments also would require that teachers become actively involved in the assessment process, while they are learning.

Objectives

The purpose of this paper is to report the development and pilot testing of a set of performance assessments, used to determine classroom teachers' measurement competencies in the areas covered in the Standards for Teacher Competence (AFT/NCME/NEA, 1990). A second purpose of the study is to demonstrate how the use of performance assessments in a graduate-level, classroom-assessment course can enhance students' learning and increase the likelihood that the measurement concepts presented will carry through to the classroom.

Methods

Twenty-nine graduate students, enrolled in a course in educational measurement, were pre-tested and post-tested using the Teacher Competencies Assessment Questionnaire (Plake & Impara, 1992). This course was required for completion of the requirements for their master's degree. For most of the students, this was the final course in their degree program. The course was taught during summer school over a six week period. The class met for two and a half hours, three times per week

The syllabus had been revised so as to incorporate more classroom assessment activities. In addition a performance assessment structure was added, mindful of the Standards for Teacher Competence in Educational

Assessment of Students (AFT, NCME, & NEA, 1990); eight performance-based tasks were created for the course. Completed tasks were used to determine achievement of the course objectives and were used in lieu of traditional multiple-choice, open-response, and essay examination assessments. The tasks were designed so that the students worked individually in some cases and in small groups for other. An expanded description of the Tasks is appended to this paper.

The instructor used a lecture format for the first half of class (60-90 minutes), supplemented with readings from Educational and Psychological Measurement and Evaluation (Hopkins, Stanley & Hopkins, 1990). Students usually worked on the nine performance-based tasks during the second half of class. Approximately a third of the students were experienced classroom teachers working on a master's degree and two-thirds were students in a combined teacher certification/master's degree program who had just completed student teaching. To determine the entry-level assessment skills of the students, the Teacher Competencies Assessment Questionnaire (Plake & Impara, 1992) was administered.

Table 1 matches the Standards with the nine performance-based tasks developed for the course. To guide students in developing the skills associated with choosing assessment methods (Standard 1), the first task required each student to identify three standardized tests applicable to his/her area of interest and read critical reviews of the instruments (e.g. Mental Measurement Yearbook). Students wrote summaries addressing issues of reliability and validity about each test, determined the test most suitable for their purpose, and wrote to the publishers requesting an examination copy (copies of the requests were given to the instructor).

Table 1: Standards Assessed by Questionnaire and Tasks

TASK 1 - Identifying Published Measurements
TASK 2 - Developing An Objective Test
TASK 3 - Summarizing Objective Test Results
TASK 4 - Conducting an Item Analysis
TASK 5 - Validity & Reliability of Objective Tests
TASK 6 - Group Standardized Test Project
TASK 7 - Developing a Performance Assessment
TASK 8 - Essay Item

Standards for Teacher Competence in Educational Assessment of Students	**Performance Assessment Tasks**
Standard 1: Teachers should be skilled in *choosing* assessment methods appropriate for instructional decisions	TASK 1
Standard 2: Teachers should be skilled in *developing* assessment methods appropriate for instructional decisions	TASKS 2, 3, 4, 5, & 7
Standard 3: The teacher should be skilled in administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods.	TASKS 3 & 6
Standard 4: Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.	TASKS 6 & 7
Standard 5: Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments	TASK 6 & 7
Standard 6: Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.	TASK 6 & 7
Standard 7: Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.	TASK 8

Tasks 2, 3, 4, 5, and 7 were designed to strengthen students' skills in developing assessment methods needed for making instructional decisions (Standard 2). The first set of activities involved the construction of a 10-item, multiple-choice test. Working in groups of two or three, students identified purposes of the instrument, developed items, and pilot-tested the instrument.

As a culminating activity for the descriptive statistics component of the course, each student summarized the results of the pilot-test, using measures of central tendency and z-scores. After completing the summarization task, the group calculated difficulty and discrimination values for the items, and the results of this analysis were used to revise the first draft of the test. Task 5, completed individually, required the students to discuss ways to establish the validity and reliability of their instruments and to calculate reliability coefficients for their instruments.

Task 7 also assisted students in developing the assessment methods needed for making instructional decisions. In this task each student identified a content area for developing a performance assessment activity, developed the activity, and created a scoring rubric for grading the product. Students also wrote about how the information from the activity would be used and how changes would be made if the activity were a part of end-of-course testing.

To gain skills in administering, scoring, and interpreting assessment instruments (Standard 3), the students completed Task 3 and Task 6. For Task 3, students were asked to administer, score, and interpret the multiple choice test they had constructed. In Task 6, a small group of two or three students, self-administered a battery of standardized tests, scored them, and created a composite character by combining results from individual tests. The group reported the results, describing the fictional character from the assessment data collected.

Tasks 6 and 7 were used to develop the students' competencies in utilizing assessment results for making decisions (Standard 4). For both Tasks 6 and 7, students were required to use the assessment information they had gathered to support an instructional decision. Tasks 6 and 7 were also used in

developing competencies relating to valid grading measures (Standard 5) and communicating results of assessments to stakeholders (Standard 6).

To develop teacher ability to recognize unethical use of assessment methods and information (Standard 7), Task 8 was modified. In this activity, the class discussed the ethical implications of the use of the assessment in various scenarios. To begin the discussion, a list of "questionable" assessment practices, collected from textbooks, teachers' accounts, and newspaper reports, were presented to the class. In small groups, the students discussed each practice and indicated whether they felt the practice to be ethical or unethical. Later, each group presented their reasoning and the "right" answer was discussed (the right answer being reflected in the outcomes of a court decision or as related by authors of the textbook).

Points were assigned to each of the tasks and total score determined the course grade. Points for tasks and final course grades were assigned as follow:

Points Assigned to Tasks

Task 1 - Identifying Published Measurements	5 points
Task 2 - Developing An Objective Test	5 points
Task 3 - Summarizing Objective Test Results	20 points
Task 4 - Conducting an Item Analysis	20 points
Task 5 - Validity & Reliability of Objective Tests	20 points
Task 6 - Group Standardized Test Project	10 points
Task 7 - Developing a Performance Assessment	10 points
Task 8 - Essay Item	10 points

Grade Scale Used for Final Grade

A+ = 95+; A = 85—94; B+ = 80—84; B = 75—79; C = 65—74; F = less than 65

During the last session of the course, the students completed an evaluation of the course, and the Teacher Competencies Assessment

Questionnaire (Plake & Impara, 1992) was readministered. The students' pre and post test scores were compared by use of the dependent t-test. In addition, students' total performance task score, reflected in the course grades, were to be correlated with their post-test scores on the Assessment Questionnaire to estimate concurrent validity between the two measures. Due to the lack of variability on students' end-of-course grades based on their Task performance, the principal investigator created an instrument to measure students' perceptions of competency relating to each of the eight Tasks. The Classroom Assessment Tasks contains 37 items distributed among the eight Task sub-heading. Respondents are asked to select their degree of competence on a seven-point scale (0=Not Competent to 6=Very Competent). The instrument is appended to this paper.

The instrument was mailed to the students six months after taking the course. At the same time the Classroom Assessment Tasks was administered to a group of students who were beginning a graduate-level measurement course at the same university. These beginning students also completed the Assessment Questionnaire. Scores on the Assessment Questionnaire for the beginning students were compared with pre-test scores of the summer students to determine initial group equivalence. Then scores on the Classroom Assessment Tasks for the two groups were compared to probe group differences in reported competencies.

Results

From the time they began the course to when they completed it, students in the summer measurement course significantly increased their scores on the Assessment Questionnaire. Table 2 shows that students in the class had an average pre-test score of 24.2 correct answers out of 35 questions and a post-

test score of 27.3 ($t = 6.31$, $p < .000$). Plake, Impara, and Fager (1992) reported a mean of 23.2 and standard deviation of 3.33 for the national sample of 555 teachers whose educational background was more varied than the study's sample.

Table 2.

Pre-Test and Post-Test Results for Students on the Assessment Questionnaire

	n	X	s	t
Pre-Test	29	24.2	2.8	
Post-Test	29	27.3	3.0	6.31***
National Sample	555	23.2	3.3	

*** p<.000

Comparisons of students' results on the end-of-course administration of the Assessment Questionnaire, with their grades in the course as a summary of total points for Tasks completed, could not be done due to a lack of appreciable variation in their grades; the class average grade was 96.4. A t-test between the beginning students' and the summer students' pre-test scores on the Assessment Questionnaire revealed no significant initial differences on this measure.

An investigation of the measurement properties of the Classroom Assessment Tasks instrument found a reliability estimate (Cronbach's alpha) of .99, and an accompanying factor analysis of the Classroom Assessment Tasks data yielded a one-factor solution that accounted for 78% of the variance.

Classroom Assessment Tasks data were gathered from 22 of the 29 students six months after completing the course and 29 graduate students beginning the same course in measurement. A comparison of the two groups on the Classroom Assessment Tasks indicated substantial significant differences in their perceived competencies in completing the eight assessment tasks. Table 3 shows that students who completed the course reported an average of 193.6 scale points (0 = minimum, 222= maximum) while beginning students reported an average of 82.3 points on the same scale.

Examination of course evaluations from the summer school students revealed that all of the students (n=25) felt the course objectives were met. Interest in the course was rated high by 11 students, average by 14 students, with none of the students reporting low interest. In response to open-ended questions about the strong points of the course (both content and teaching), 13 students mentioned the performance-based tasks as strengths of the course and 9 of the students mentioned the relevance of the course to their teaching. One student felt that the course over-emphasized classroom-related information, and another student felt the material needed more emphasis on classroom implementation.

Table 3.

Comparison of Students Six-Months after Course Completion and Beginning Measurement Students on the Classroom Assessment Tasks

	n	X	s	t
Students Six-Month Follow-Up	22	193.6	27.0	
Beginning Measurement Students	29	82.3	48.621	10.39***

*** p<.000

Conclusions

While very preliminary in nature, the results of the study indicate that students who complete a measurement course that is performance based do increase their assessment competency as measured by the Assessment Questionnaire. Further, the study revealed that although there were no appreciable initial differences on the Assessment Questionnaire between two groups of graduate students as they began their measurement coursework,.beginning students assessed their competencies at measurement tasks well below those who had completed a measurement course.

Perhaps performance-based educational measurement courses, in stressing higher order thinking skills, actively involving students in the learning process, and personalizing instruction, increase the relevance of assessment to practitioners. And, in turn, will increase the likelihood that the measurement concepts presented will carry through to the classroom.

References

Airasian, P. W. (1991). Perspectives on measurement instruction. Educational measurement: Issues and practice, 10 (1), 13-16, 26.

American Federation of Teachers, National Council of Measurement in Education, & National Education Association. (AFT/NCME/NEA, 1990) Standards for teacher competence in educational assessment of students. Educational Measurement: Issues and Practice, 9 (4), 30-32.

Frechtling, J. A. (1991). Performance Assessment: Moonstruck or the real thing? Educational measurement: Issues and practice, 10 (4), 23-25.

Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). Educational and psychological measurement and evaluation (7th ed.) Englewood Cliffs, New Jersey: Prentice Hall.

Jett, D. L., & Schafer, W. D. (1992, April). Classroom teachers move to center stage in the assessment arena -- ready or not!. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20 (8),15-21

Marso, R. N., & Pigge, R. L. (1992, April). A summary of published research: Classroom Teachers' knowledge and skills related to the development and use of teacher-made tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Mehrens, W. A. (1992). Using performance assessment for accountability purposes. Educational measurement: Issues and practice, 11 (1), 3-9.

Moss, P. A., Beck, J. S., Ebbs, C., Matson, B., Muchmore, J., Steele, D., Taylor, C., & Herter, R. (1992). Portfolios, accountability and an interpretive approach to validity. Educational measurement: Issues and practice, 11 (3), 12-21.

O'Neil, J. (1992). Putting performance assessment to the test. Educational leadership, 49 (8), 14-19).

O'Sullivan, R. G. (1992). Classroom assessment tasks. Greensboro, NC: School of Education, University of North Carolina at Greensboro.

O'Sullivan, R. G., & Chalnick, M. K. (1991). Measurement related coursework requirements for teacher certification and recertification. Educational Measurement: Issues and Practice, 10, 17-19, 23.

O'Sullivan, R. G., McColskey, W., & Harman, P. (1992, March). Implementing performance assessment in middle school science programs. Paper presented at the annual meeting of the Eastern Educational Research Association, Hilton Head, SC.

Plake, B. S., & Impara, J. C. (1992). Teacher Competencies Questionnaire description. Lincoln, NE: University of Nebraska.

Plake, B. S., Impara, J. C., & Fager, J. J. (April 1992). Assessment competencies of teachers: A national survey. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco, CA.

Plake, B. S. , Impara, J. C., & Sanders, J. R. (April 1992). Report of the Task Force on the Development of Standards for Administrators in the Assessment of Students: A Kellogg grant project involving NCME, AASA, NAESP, and NASSP. Symposium presented at the annual meeting of the National Council of Measurement in Education, San Francisco, CA.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38, (3), 57-62.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments political rhetoric and measurement reality. Educational Researcher, 21 (4), 22-27.

Stiggins, R. J. (1991a). Relevant classroom assessment training for teachers. Educational measurement: Issues and practice, 10 (1), 7-11.

Stiggins, R. J. (1991b). Assessment literacy. Phi Delta Kappan, 72, 534-539.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 70, 703-713.

Wiggins, G. (1992). Creating tests worth taking. Educational leadership, 49 (8), 26-33

Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. Journal of Teacher Education, 42 (1), 37-42.

Wolf, E. P., LeMahieu, P. G., & Eresh, J. (1992). Good measure: Assessment as a tool for educational reform. Educational leadership, 49 (8), 8-13.

ASSESSMENT TASKS

Task 1 (Individual): Identifying Published Test Materials
Identify topic area for assessment
Find three instruments that assess the chosen topic area. (At least two must be reviewed in one of the Mental Measurement Yearbooks or some other source)
Write a short summary of each of the tests
Select the test you feel is most appropriate to your purpose(s) and explain why you selected it.
Write to the publisher of the test and request information about the test.

Task 2 (Partners): Developing an Objective Test
Identify the purpose of the measurement
Specify the goals and objectives of the instrument.
Develop a test blue-print of specifications.
Draft 10 test items (5 items each)
Have your partner review the items
Pilot test the items with 20 individuals.

Task 3 (Individual): Summarizing the Results of an Objective Tests
Calculate measures of central tendency and variability for test
Include z-scores and percentiles

Task 4 (Partners): Conducting an Item Analysis
Calculate *p*-values for items
Calculate *D*-values for items
Report findings in tabular form (a great example is in the text))
Highlight items that are questionable
Eliminate questionable items from item analysis
Write new items for ones that you eliminated
Revise items that are salvageable
Revise specifications blueprint and submit with new test

Task 5 (Individual): Validity & Reliability of Objective Tests
Discuss how you would establish the validity of your instrument.
Calculate a split-half reliability coefficient and adjust with the Spearman-Brown Prophesy formula.
Calculate a KR-20 reliability coefficient

Task 6 - (Group): Standardized Test Project
Working in groups of 2-3, each person self-administers two published, standardized tests that are available in the test battery packet. Students will use the technical manuals that accompany each instrument to find out how to administer, take, score, and interpret each test. Students will also have to look at the technical merit of each test as it is reported in the technical manual (validity and reliability information). Students will then combine the results of their individual tests in a hypothetical career counseling situation. These combined efforts will take the form of a report.

Task 7 - Developing a Performance Assessment
Select curriculum area that lends itself to performance assessment.
Develop a performance assessment item, identifying the curriculum areas to be assessed by the item
Determine how you will "grade" the item. What rubric will you use.
How will the results be used? What classroom decisions might be made based on this information?
How might you change the item, if it were to be used as part of end of course testing? Why?

Task 8 - Ethical Issues (Take-home final)
Distribute description of situation that requires a measurement decision that reflects ethical values.
Write essay to be turned in on Exam Day
Discuss positions in class during final class meeting

EDUCATION 670: Course Outline
SUMMER 1992
EDUCATIONAL MEASUREMENT & EVALUATION

Robert L. Johnson
Office: 210 Curry
Phone: 334-5883

Mon. Tues. & Thurs., 2-4:30 p.m.
Room: 247 Curry

Office Hours: Before class or by arrangement

Required Course Text :

Hopkins, K.D., Stanley, J.C., & Hopkins, B. R. (1990). Educational and Psychological Measurement and Evaluation, Seventh Edition. Englewood Cliffs, N.J.: Prentice Hall.

Course Schedule

Dates	Lecture Topic	Reading
7/2	First Meeting/Orientation Introduction - Measurement & Evaluation Types of Educational Assessment	3 - 19
7/6	No Class - University Holiday	
7/7	Jackson Library Meeting - Task 1 Identifying Published Measurements	
7/9 - 7/16	Descriptive Statistics (5 hours) Task 2 - Developing An Objective Test Task 3 - Summarizing Objective Test Results	20 - 75 165-192; 224-266
7/20	Improving the Quality of Test Items Task 4 - Conducting an Item Analysis Task 5 - Revising Items	267-285
7/21 -7/23	Validity & Reliability of Classroom Assessments Establishing Test-Retest Reliability Task 6 - Validity & Reliability of Objective Tests	76 - 164
7/27 - 7/28	Standardized Tests Edwards Personal Preference Scale Task 7 - Group Standardized Test Project	341 - 470
7/30 - 8/6	Alternative Assessment Task 8 - Developing a Performance Assessment	193-223
8/10	Ethical Issues (receive take-home essay item)	320 -338
8/11	Final Examination Task 9 - Essay Item	

Course Objectives:

This course is designed to develop an informed consumer of standardized tests as well as improve test construction skills. The course will cover:

- Basic principles of measurement and evaluation;
- Methods of scoring and interpreting tests;
- Construction and use of teacher-made tests;
- Statistical concepts basic to understanding and interpreting test data.

Assumptions and Procedures

Students taking Education 670 have substantial individual differences in background ranging from no statistics, measurement, or math in recent years to those with a rather extensive background. This course will be taught on the basis that students enter the course with a minimal math background.

An initial assumption made is that graduate students do not require that all text material be discussed in a lecture. A second assumption is that certain of the topics will require more detailed treatment than will others because of their technical nature. It is expected that all students will read the text before the class meeting and be prepared to ask questions and be asked questions based on their prior readings. Certain of the topics will be developed in far more detail than will others in lecture.

Since the major objective of the course is to help students acquire a basic understanding of educational measurement and evaluation in order to make informed decisions about appropriate testing as well as improve test construction, administration, scoring, and interpretation skills, the course requirements have been designed to promote practice of these skills. Each student will be asked to complete nine tasks that will provide practice in administration, scoring, interpretation, and evaluation of classroom assessment instruments. The nine tasks are performance based and intended to actively involve students in the assessment process. Each task will be described in detail so that students understand what is expected.

The evaluation of student performance will be based upon the following:

Task 1 - Identifying Published Measurements	5 points
Task 2 - Developing An Objective Test	5 points
Task 3 - Summarizing Objective Test Results	20 points
Task 4 - Conducting an Item Analysis	20 points
Task 5 - Validity & Reliability of Objective Tests	20 points
Task 6 - Group Standardized Test Project	10 points
Task 7 - Developing a Performance Assessment	10 points
Task 8 - Essay Item	10 points

Grade Scale Used for Tasks

A+ = 95+; A = 85—94; B+ = 80—84; B = 75—79; C = 65—74; F = less than 65

CLASSROOM ASSESSMENT TASKS

The tasks listed below identify classroom assessments that teachers may be called on to perform. Please indicate by circling the appropriate number below how competent you feel about performing each of the tasks.

Task 1 : Identifying Published Test Materials	Not Competent						Very Competent
1. Find three published instruments that assess a chosen educational topic area	0	1	2	3	4	5	6
2. Select the test you feel is most appropriate to your purpose(s) and explain why you selected it.	0	1	2	3	4	5	6
3. Write to the publisher of the test and request information about the test.	0	1	2	3	4	5	6
Task 2: Developing an Objective Test							
4. Identify the purpose of a test instrument	0	1	2	3	4	5	6
5. Specify the goals and objectives of the instrument.	0	1	2	3	4	5	6
6. Develop a test blue-print of specifications.	0	1	2	3	4	5	6
7. Draft test items	0	1	2	3	4	5	6
8. Review/edit the items	0	1	2	3	4	5	6
9. Pilot test the items	0	1	2	3	4	5	6
Task 3: Summarizing the Results of an Objective Tests							
10. Calculate measures of central tendency for the test	0	1	2	3	4	5	6
11. Calculate measures of variability for the test	0	1	2	3	4	5	6
12. Calculate z-scores for the test	0	1	2	3	4	5	6
13. Calculate percentile equivalents for the test	0	1	2	3	4	5	6
Task 4: Conducting an Item Analysis							
14. Calculate item difficulty values	0	1	2	3	4	5	6
15. Calculate item discrimination values	0	1	2	3	4	5	6
16. Report findings in tabular form	0	1	2	3	4	5	6
17. Identify items that are questionable based on the item analysis information	0	1	2	3	4	5	6
18. Revise items based on the item analysis information	0	1	2	3	4	5	6
19. Revise specifications blueprint based on the item analysis information	0	1	2	3	4	5	6
Task 5: Determining the Validity & Reliability of Objective Tests							
20. Determine how to establish the validity of your instrument.	0	1	2	3	4	5	6
21. Determine how to establish the reliability of your instrument	0	1	2	3	4	5	6
22 Calculate a reliability coefficient.	0	1	2	3	4	5	6
23. Improve the reliability of your instrument	0	1	2	3	4	5	6

Task 6: Using Standardized Tests	Not Competent						Very Competent
24. Identify the validity of a standardized test from the technical manual	0	1	2	3	4	5	6
24. Identify the reliability of a standardized test from the technical manual	0	1	2	3	4	5	6
25. Administer a standardized test using the technical manual instructions	0	1	2	3	4	5	6
26. Score a standardized test using the technical manual instructions	0	1	2	3	4	5	6
27. Interpret the results of a standardized test using the technical manual instructions	0	1	2	3	4	5	6
28. Generate a profile for a student from a battery of standardized tests	0	1	2	3	4	5	6
Task 7: Developing a Performance Assessment							
29. Identify a curriculum area that lends itself to performance assessment.	0	1	2	3	4	5	6
30. Develop a performance assessment item, identifying the curriculum areas to be assessed by the item	0	1	2	3	4	5	6
31. Pilot-test the performance assessment item	0	1	2	3	4	5	6
32. Determine how you will score/grade the item.	0	1	2	3	4	5	6
33. Determine what classroom decisions might be made based on the performance assessment information?	0	1	2	3	4	5	6
34. Modify a performance assessment based on the decision to be made from the assessment	0	1	2	3	4	5	6
Task 8: Ethical Issues							
35. Identify measurement decisions that reflect ethical values	0	1	2	3	4	5	6
36. Judge the ethical merits of testing decisions	0	1	2	3	4	5	6
37. Evaluate measurement instruments using the Standards for Educational and Psychological Testing	0	1	2	3	4	5	6